ANNUAL PROGRESS REPORT

05/01/2006 - 04/30/2007

NASA 05-AISR05-0143

Bayesian Source Separation for Astrophysical Spectra: Application to PAHS

Kevin H. Knuth (PI)
kknuth@albany.edu (518) 442-4653
Department of Physics, PH 211, University at Albany, Albany NY 12222
http://knuthlab.rit.albany.edu/

General Summary

We propose to apply Bayesian source separation techniques to a very important subset of astrophysical spectra, the emission spectra of PAHs. The PAH species, ubiquitous in our Galaxy and in external galaxies, are one of the most important molecular species in the pathway to life. Their spectra are the result of the blending of the complex spectra of individual PAH species, possibly at different temperatures, and to date they have defied all attempts to unscramble them into individual contributors. We propose to apply Bayesian source separation for the first time in order to quantitatively determine which of the many hundreds of individual PAH species contribute most significantly to the combined emission spectra of PAHs. Our research will lead to powerful new tools for decomposing, analyzing and characterizing spectra of these critical species. Furthermore, these tools will be readily generalizable to a wide array of spectral analysis problems ranging across astrophysics, Earth science, and biology.

Summary of Our Efforts

Our first efforts focused on studying the behavior of the traditional non-negative least squares technique applied to the problem of identifying the PAHs present in an artificial mixture of PAH spectra. We showed the problem to be easily solved by non-negative least-squares in the case of a handful of PAHs. Our expectations were that this would break down with the number of PAH species exceeding 25 or so. However, we found that the technique worked well for mixtures involving nearly 80-100 PAHs. Our other key prediction did hold, and that was that the non-negative least-squares technique always resulted in a contribution for every PAH, even those that were not present in the data set for which it returned a small, albeit finite contribution. As we also predicted, there were several species of PAHs that we found that were difficult to separate.

During this phase, the Ames team (Allamandola, Bauschlicher, Cami and Peeters) worked to build their database of PAH spectra, which currently holds the spectra of almost one thousand PAH species. This database will be essential for our future algorithms which will test for the presence of these PAH species in real datasets.

The next phase consisted of dealing with the characterization of the background spectrum, which typically consists of one or more blackbody radiators in addition to other unknown effects. To solve this problem, we utilized a new Markov chain Monte Carlo algorithm called Nested Sampling that allows us to explore spectral models with varying

numbers of components. Two algorithms were developed. One algorithm models the background as a sum of blackbody radiators at distinct temperatures. This is useful when the infrared signal emanates from one or more optically thin molecular clouds lying along the same line-of-sight. The second algorithm models the background spectrum as a sum of Gaussians. While not physically motivated like the Planck blackbody, the sum of Gaussians model can capture spectral features of unknown origin. These two algorithms have now been combined using Nested Sampling. The result is an excellent algorithm that describes the blackbody radiation in terms of both blackbody radiators and unknown sources. We are currently polishing this algorithm for submission into the AISRP Code and Algorithm Library.

We are now working toward uniting the background spectrum model with the PAH spectrum model to simultaneously learn the PAH contributions and the background properties. To this end, three members of Knuth's team at the University at Albany are working to prepare Knuthlab's 16-node Beowulf cluster for this task.

Personnel, Progress Description, and Summary of Accomplishments

In the summer of 2006, we hired two physics graduate students Man Kit Tse and Phil Erner to work on developing MatLab code for PAH spectrum classification. Man Kit Tse continued to work with me throughout the year on the project and was instrumental in developing the background estimation algorithm based on a Mixture of Gaussians model. We have been working to hire Deniz Gencaga, and expert in source separation applied to astrophysical spectra, as a postdoctoral fellow, but there have been visa delays as we describe below.

Several physics/computer science undergraduate students have volunteered for this project for Independent Study. Haley Maunu is currently working on background spectrum estimation using a spline model. Joshua Choinsky worked to develop the mixture of Planck Blackbodies algorithm. Ben Rosenblum, Christopher Stiles, and Brian Nathan are developing code for use on our laboratory's 16-node Beowulf cluster.

Duane Carbon at NASA Ames researched in detail the behavior of the non-negative least-squares algorithm, and worked with the PAH group to build the PAH database that is central to our efforts.

Kevin Knuth has worked to oversee this process as well as guiding the students on the developments using the Nested Sampling algorithm and integrating the algorithmic results across the team. He has also given several talks on the topic as well as writing abstracts and papers with Duane Carbon. He has worked also to integrate these efforts with past efforts in characterizing the environments of planetary nebulae which are the sources of many of the PAH spectra we are studying. In the process of our work here, we have further developed a Bayesian histogramming technique that we are currently polishing up for submission into the AISRP Code and Algorithm Library. The paper describing this technique will be submitted to *Phys Rev E*.

Schedule Status, Delays/Problems Experienced, and Corrective Actions Plan

We have been working to hire Deniz Gencaga as a postdoctoral fellow with the aim of him becoming a key member of this team. Deniz is an expert in source separation applied to astrophysical spectra, and is about one in ten people in the world with a skill set so uniquely qualified for this project. Deniz was to graduate in January 2007 and join us by February. However, his visa requirements dictated that he must have his Bachelor's Degree in hand before we could begin the hiring process. This unfortunately has been delayed due to the fact that his university's requirement of graduation is the publication or a peer-reviewed paper. Deniz's paper was held up in review for over 9 months delaying his actual graduation date, and delaying his joining our group. This is now in the process of resolution, and we expect that Deniz will join us in early May 2007.

Due to this setback, we will be concluding our study of mixtures on the order of 1000 PAHs in June 2007, and expect to have conclusions regarding PAH identifiability before the end of summer 2007.

Work Plan for next Reporting Period

For the remainder of this year and into the summer of 2007, we will be working to integrate the background spectrum estimation code with the PAH estimation code and testing it on both synthetic mixtures and real data acquired by the Infrared Space Observatory (ISO). We will also work to identify classes of PAHs that are difficult to separate (some of this has already been performed by Duane Carbon). As described above, we will be concluding our study of mixtures on the order of 1000 PAHs in June 2007, and expect to have conclusions regarding PAH identifiability before the end of summer 2007. As we move into the next year, with the success of our techniques, we will begin to deal with more subtle problems such as sets of PAHs at distinct temperatures.

Publications and Presentations

1. Knuth K.H., Tse M.K., Choinsky J., Carbon D.F. 2007. Bayesian source separation applied to identifying complex organic molecules in space, **submitted** to *Statistical Signal Processing Workshop 2007*, Madison WI, June 2007.
2. Carbon D.F., Tse M.K., and Knuth K.H. 2007. Bayesian source separation for PAH spectra, *209th Meeting of the American Astronomical Society*, Seattle WA, Jan 2007.
3. Knuth K.H., Carbon D.F. 2006. Bayesian Source Separation for Astrophysical Spectra: Application to PAHs, *NASA Applied Information Systems Research Program Workshop*, Adelphi MD, Oct 2006.
4. Knuth K.H. Optimal data-based binning for histograms. **To be submitted** to *Phys Rev E*